3.5.1 STANDARD DATA OUTPUT (SDO) FILE

Introduction

The Standard Data Output (SDO) file is an ASCII file that is used as input to the StEPS Batch Update program. This SDO file eliminates the need for specific update programs for different surveys and/or collection activities. Whenever a collection activity (e.g., check-in, keying) takes place, the output must be placed in the SDO format before the information can be applied to StEPS. This is true for any batch update file being loaded into StEPS. The batch update program will then apply the data from this SDO to the respective survey data sets.

The purpose of the SDO is to:

- Apply *collected* data to the Item, Notes and/or Respondent Text files
- Update *specific* fields on the Control, Item, Notes and/or Respondent Text files
- Add *new* ID's and related information to the Master and Stat Period Control files
- Generate status change actions (delete, ghost, merge, restore, reserve) to *existing* IDs in the Master and Stat Period Control files

SDO file names will vary depending on the source of the file. For example, SDO files created as a result of keying in NPC are placed in the Master Control System (MCS) using TMO's naming conventions. Other sources for SDO files include sampling, CSAQ, TDE, mass correction, etc. Each of these files will be named differently. Naming requirements for SDO files are given in the respective source documentation.

File Layout

The number of records included in an SDO file is not limited, however the length of each record cannot exceed 229 characters.

The first 49 characters on all SDO records will have the **SAME** layout (with the exception of the *Identification Number* field on the mail group records). Characters 50 and beyond will vary depending on the "record type" specified in position 31, as described in the following sections:

3.5.1.1 SDO CHARACTERS 1-49

NOTE: The first 49 characters on all SDO records will have the same layout (with the exception of the *Identification Number* field on the mail group records). Characters 50 and beyond will vary depending on the "record type" specified in position 31. These tables give the Field Name and, in capital letters, the VARIABLE NAME used by the SDO file and batch update program.

Position	Length	Field Name	Description
1-6	6	Survey Identifier (SURVEY)	See StEPS Planning/Decision Document #3 for possible values. This survey identifier will be imprinted on the StEPS standard label. There is also a 1-1 mapping of this identifier to the 3-digit numeric StEPS check-in code used in the bar code.
7-12	6	Statistical Period (STATP)	For monthlies: YYYYMM, where 'YYYY' denotes the survey year and 'MM' denotes the survey month; Else YYYYfp where 'YYYY' denotes the survey year, 'f' denotes the frequency, and 'p' the period within the year. Quarterlies 1995Q3 = Quarter 3 1995 Biannuals 1995B2 = Half 2 1995 Annuals 1995A1 = 1995 Annual supplements 1995S1 = 1995 Biennials 1995N1 = 1995 Triannuals 1995T1 = 1995 Quinquennials 1995U1 = 1995
13-28	16	Identification Number (ID)	See StEPS Planning/Decision Document #4 for explanation of the StEPS ID field. (For many surveys this will simply be 11 digits, as on the bar code, followed by 5 blank characters. The additional 5 characters are sometimes needed, however, to handle more complex reporting arrangements in some surveys (e.g., children records). NOTE: For the mail group records, this field will *instead* contain an 8-character value comprised of alpha (characters 1-6) and mail group (characters 7-8). This 8-character value will be left-justified, blank filled, in the Identification Number field on the SDO record.

Position	Length	Field Name	Description
29-30	2	Source (SOURCE)	Indicates the source of the data output record. Often the record is from a data collection method and this field indicates the source of the data response. The first character of source will be placed in the Data Source (DTSRCE) field in the StEPS Stat Period Control file. **Data Collection Source Fields:** *('_' in the 2nd character represents <blank>)* F_ Report form data captured via keying C _ CATI E_ EDI P_ CAPI T_ TDE I_ Internet Q_ CSAQ A_ Analyst phone contact J_ Clerk phone contact R_ Respondent phone in X_ Fax S_ Self filer O_ Other NOTE: The batch update program will ONLY update the DTSRCE field on the Stat Period Control file if there is an 'entire form' update for the current relative stat period (i.e., Action = 'R', Record type = 'I', and Relative stat period = '00'). **Other Source Fields:** CK Bar code check-in SM Sampling FF Future file OS Other source used to obtain 'other' items; this code MUST be used only for items that have a value of 'S' in the ISOURCE field in the Item Data Dictionary. MS Miscellaneous source of non-collection batch update records

Position	Length	Field Name	Description
31	1	Record Type (TYPE)	Indicates the type of data output record. Different types will have different record layouts in *characters 50 and beyond* and the batch update program will perform different *actions* based on this record type. Note: These various actions also depend on the "Record action" (found in position 32) and "Source" fields. Record types include the following: R Respondent text C Control file I Item data N Notes file H Collection history file M Mail group file S Item data with corresponding data flag to be updated T Roster item data U Roster item data with corresponding data flag to be updated

Position	Length	Field Name	Description
32	1	Record Action (ACTION)	Indicates the action to be taken for this data output record. Possible values include the following: R Replace Replace fields (on the appropriate files) with the data supplied in the SDO record. This is the most common record action coming from keying for a base stat period. It indicates that, except for certain items specified in the batch update spec (these exceptions are items indicated in the data dictionaries with ISOURCE of 'S' or FWDFLG of 'Y'), the items for the given ID/stat period are to be replaced (or added) with items in this transmission. Rules for what to do when an ID has already received items for this stat period are specified in the batch update spec. C Change Change (or add) values (on the appropriate files) for those fields specified in the SDO only. This type of action is useful for changes to just a few items or control fields when you wish the other items and fields to be left untouched. This is the action that will be on the file for corrections to prior stat periods (i.e., Stat period is current stat period, Action = 'C'; Relative data period indicates which stat period file should be changed). L Link Change fields (on the appropriate files) with the new values specified in the SDO record, but also **link** to other records and make changes as well. For example, if the record TYPE = 'M' (for mail group) and the ACTION = 'L', this means to not only change the fields indicated on the mail group file, but also to apply these changes on the Master Control file for ALL IDs with the *same alpha and mail group code* as located in the Identification Number (ID) field (8-character value) on the SDO record. A Add this ID D Delete this ID G Ghost this ID V Reserve this ID S Restore this ID NOTE: For actions A, D, G, V, S: Certain information will be required and other information will automatically be changed in the Batch Update program as a result of a particular action; see Status Changes Batch Update spec for more information.

Position	Length	Field Name	Description
33	2	Relative Statistical Period (RELSTP)	Indicates whether the data on this output record is for the current statistical period (as keyed in the Statistical Period field), or for a prior statistical period. It is a 2-digit relative statistical period indicator. Possible values include: 00 Current statistical period 01 Prior statistical period 02 2 statistical periods back from the current (aka, previous prior statistical period) NOTE: A relative data period value other than '00' can ONLY occur where record TYPE = 'I' (item) or 'S' (item with data flag).
35-42	8	Date (DATE)	The date on which this data output record was created. The format is YYYYMMDD. StEPS will enter this date into the DKDTE field on the Stat Period Control file automatically for records where SOURCE is from one of the data collection methods (see positions 29-30, "source", above). This will only occur on 'entire form' updates where Action = 'R', TYPE = 'I', and Relative stat period = '00'.
43-46	4	Data Batch Number (BATNUM4)	For cases keyed at NPC, this field contains the batch number in which the form was placed for keying (aka lot, work unit). StEPS will automatically enter this number into the DBATCH field on the Stat Period Control file. This will only occur on the 'entire form' updates where Action = 'R', TYPE = 'I', and Relative stat period = '00'. (This field is zero-filled for all other situations).
47-49	3	Sequence Number (SEQNUM)	For cases keyed at NPC, this field contains the sequence number of the form in the batch. Therefore, *all records for a given questionnaire* will have the same sequence number on the output records. StEPS will automatically enter this number into the DSQNUM field on the Stat Period Control file. This will only occur on the 'entire form' updates where Action = 'R', TYPE = 'I', and Relative stat period = '00'. (This field is used for sorting in the batch update program and should be zero-filled for all other situations).

3.5.1.2 CHARACTERS 50 and BEYOND

NOTE: Characters 50 and beyond of the SDO file have layouts specific to their respective record TYPE (see position 31 in section 3.5.1.1).

Record Type	Field Description
C (Control file)	A series of 6-character mnemonic fields followed by a 36 character value. There can be a maximum of 4 mnemonic/value sets on any one 'C' record. **Mnemonic** A 6-character field name from the StEPS Master or Stat Period Control file for any of its alphanumeric fields (if less than 6 characters, left justify and then enter blanks to fill the entire 6-character space) **Value** A 36-character left-justified, blank-filled value for corresponding mnemonic. NOTE: Any 'date' in the Master or Stat Period Control file fields may be entered in either YYYYMMDD or YYMMDD format. Numeric data may be entered here as character and the batch update program will load correctly (eg. 999.99) This data will be applied to the appropriate StEPS Control file (Master or Stat Period) for the respective ID/stat period).
I (Item file)	A series of 5-character key codes followed by a 13-digit value. There can be a maximum of 10 key code/value sets on any one 'I' record. **Key code** A 5-character field that is in the StEPS Item Data Dictionary. There is a 1-1 relationship between this key code and the StEPS item code and often they will be the same. (If the key code is less than 5 characters, left-justify and then enter blanks to fill the entire 5-character space). NOTE: It is imperative that the key code in the Item Data Dictionary exactly match what is output from keying and placed in the key code space on the SDO records. **Value** A 13-digit right-justified value for the corresponding key code. This value may begin with a '-' sign. NOTE: This data will be applied to the StEPS Item data file for the respective ID/stat period.

Record Type	Field Description
R (Respondent Text file)	A 5-character key code followed by a 160-character value. There can be only 1 key code/value set on any one 'R' record. **Key code** A 5-character field that is in the StEPS Respondent Text Data Dictionary. If the key code is less than 5 characters, left- justify and then enter blanks to fill the entire 5-character space. NOTE: It is imperative that the key code in the Respondent Text Data Dictionary exactly match what is output from keying and placed in the key code space on the SDO records. **Value** A 160-character left-justified value
N (Notes file)	A 5-character key codes followed by a 160-character value. There can be only 1 key code/value set on any one 'N' record. **Key code** A 5-character field that is identified in a specific keying spec (could simply be sequential number). If the key code is less than 5 characters, left-justify and then enter blanks to fill the entire 5-character space. **Value** A 160-character left-justified value
H (Collection History file)	There can be only 1 collection code/value set on any one 'H' record. **Collection code** A 3-character field that is specified in the StEPS Decision Document # 7, which indicates different types of collection activities, including UAA's, initial mailout, follow-up, correspondence, check-in, etc. (All 3 characters are to be filled.) **Collection date** An 8-character field (in YYYYMMDD format) to represent the collection date that the collection activity (represented by the collection code) took place. **Value** A 160-character left-justified value; this may or may not be filled in, depending upon the value of the ACTION code (position 32). NOTE: Check-in actions and extension dates must be entered on record TYPE = 'C' and the batch update program will automatically generate the needed 'H' record type.

Record Type	Field Description
M (Mail Group file)	A series of 6-character key codes followed by a 36-character value. There can be a maximum of 4 character mnemonic/value sets on any one 'M' record. **Mnemonic** A 6-character field name from the StEPS Mail Group file for any of its alphanumeric fields. The fields on the Mail Group file are a subset of those fields on the Control files. (If the field name is less than 6-characters, left-justify and then enter blanks to fill the entire 6-character space) **Value** A 36-character left-justified, blank-filled value for the corresponding mnemonic. This data will be applied to the appropriate StEPS Mail Group file for the respective alpha/mail group.
S (Other item data with corresponding data flag)	A series of 5-character item codes (not key codes) followed by 14 characters, where the first 13 characters represent the data value of the item and the 14th character represents the data flag for this 'other' data item. There can be a maximum of 9 item code/value sets on any one 'S' record. **Item code** A 5-character field that is in the StEPS Item Data Dictionary. (If less than 5 characters, left-justify and then enter blanks to fill the entire 5-character space). **Value** A 13-digit right-justified value for a corresponding key code. This value may begin with a '-' sign. **Flag** A 1-character valid item data flag as specified in StEPS Decision Document #6. (This data will be applied to the StEPS item data file for the respective ID/stat period).

Record Type	Field Description
T (Roster item data)	There will be only one rkey1/rkey2/rtype on each 'T' record, followed by a maximum of 7 key code/value sets. **Rkey1** An 11-character field that is the unique "Key1" identifier for the roster item. This value will be on the Item Data Dictionary for the roster item. **Rkey2** An 11-character field that is the unique "Key2" identifier for the roster item. This value will be on the Item Data Dictionary for the roster item. **Rtype** A 2-character field that is the roster type for the roster item. This value will be on the Item Data Dictionary for the roster item. **Key code** A 5-character field that is in the StEPS Item Data Dictionary. There is a 1-1 relationship between this key code and the StEPS item code and often they will be the same. (If the key code is less than 5 characters, left-justify and then enter blanks to fill the entire 5-character space). NOTE: It is imperative that the key code in the Item Data Dictionary exactly match what is output from keying and placed in the key code space on the SDO records. **Value** A 13-digit right-justified value for the corresponding key code. This value may begin with a '-' sign. (This data will be applied to the StEPS item data file for the respective ID/stat period).

Record Type	Field Description
U (Roster item data with corresponding data flag)	(There will be only 1 rkey1/rkey2/rtype on each 'U' record followed by any number between 1-7 item code/value/flag sets). **Rkey1** An 11-character field that is the unique "Key1" identifier for the roster item. This value will be on the Item Data Dictionary for the roster item. **Rkey2** An 11-character field that is the unique "Key2" identifier for the roster item. This value will be on the Item Data Dictionary for the roster item. **Rtype** A 2-character field that is the roster type for the roster item. This value will be on the Item Data Dictionary for the roster item. **Item code** A 5-character field that is in the StEPS Item Data Dictionary. If the item code is less than 5 characters, left-justify and then enter blanks to fill the entire 5-character space). **Value** A 13-digit right-justified value for the corresponding key code. This value may begin with a '-' sign. (This data will be applied to the StEPS item data file for the respective ID/stat period). **Flag** A valid item data flag as specified in StEPS Decision Document #6. (This data will be applied to the StEPS item data file for the respective ID/stat period).

3.5.1.3 SDO LAYOUT SUMMARY

➤ **SDO CHARACTERS 1 - 49** (common to all SDO records)

Survey 1-6 6char	Statp 7-12 6char	ID 13-28 16char	Source 29-30 2char	Type 31 1char	Action 32 1char	Relstp 33-34 2char	Date 35-42 8char	Batnum4 3-46 4char	Seqnum 47-49 3char

➤ **SDO CHARACTERS 50 and BEYOND** (will vary depending on the TYPE value set in position 31)

TYPE = 'I' (Item data) (key code = 5 char, value = 13 char)

Key1 50-54	Val1 55-67	Key2 68-72	Val2 73-85	Key3 86-90	Val3 91-103	Key4 104-108	Val4 109-121	Key5 122-126	Val5 127-139
Key6 140-144	Val6 145-157	Key7 158-162	Val7 163-175	Key8 176-180	Val8 181-193	Key9 194-198	Val9 199-211	Key10 212-216	Val10 217-229

Type = 'C' (Control data) (var = 6 char, value = 36 char)

Var1 50-55	Value1 56-91	Var2 92-97	Value2 98-133	Var3 134-139	Value3 140-175	Var4 176-181	Value4 182-217

Type = 'N' (Notes data) (key code = 5 char, text = 160 char)

Key Code 50-54	Text 55-214

Type = 'R' (Respondent Text data) (key code = 5 char, text = 160 char)

Key Code 50-54	Text 55-214

Type = 'H' (Collection History data) (collect date = 8 char, text = 160 char)

Collection Code 50-52	Collection Date 53-60	Collection Text 61-220

Type = 'S' **(Item data)** **(**item code = 5 char, value=13 char, data flag = 1 char)

Item1 50-54	Val1 55-67	Flag1 68	Item2 69-73	Val2 74-86	Flag2 87	Item3 88-92	Val3 93-105	Flag3 106	Item4 107-111	Val4 112-124	Flag4 125
Item5 126-130	Val5 131-143	Flag5 144	Item6 145-149	Val6 150-162	Flag6 163	Item7 164-168	Val7 169-181	Flag7 182	Item8 183-187	Val8 188-200	Flag8 201
Item9 202-206	Val9 207-219	Flag9 220									